

블로그 연결망에서 콘텐츠 파워 유저의 파악 방안

김형준*, 임승환*, 김상욱*, 박선주**

*한양대학교 전자컴퓨터통신공학과

**연세대학교 경영학과

e-mail: kimhyungjun79@hanmail.net

A Method for Finding Contents-Power Users in Blog Networks

Hyung-Jun Kim*, Seung-Hwan Lim*, Sang-Wook Kim*, Sunju Park**

*Department of Electronics and Computer Engineering, Hanyang University

**Department of Business, Yonsei University

요 약

블로그 연결망 내에는 다른 사용자가 블로그 서비스를 활발하게 이용하도록 유도하는 역할을 하고 있는 특별한 사용자가 존재한다. 본 논문에서는 이러한 사용자들 가운데 콘텐츠의 영향력이 큰 사용자들을 콘텐츠 파워 유저라고 정의한다. 본 논문에서는 블로그 서비스 활성화를 위한 정책을 수립할 수 있도록 콘텐츠 파워 유저들을 파악하는 기법에 대하여 논의한다. 우선, 각 사용자가 소유하고 있는 각 게시글들의 콘텐츠 영향력을 계량화하는 방법을 제안하고, 이 값들의 합을 이용하여 해당 사용자의 콘텐츠 영향력을 계산하는 방법을 제안한다.

1. 서론

본 논문에서는 온라인 사회연결망 내에서 일반 사용자와 차별되는 고유의 영향력을 가진 사용자를 파워 유저(power user)라고 정의한다. 기존의 연구에서는 사회연결망의 위상 구조적인 특징만을 고려하여 파워 유저를 파악한 바 있다[1][2][3]. 만일, 사회연결망을 분석하는데 있어서 사용자들간의 관계의 정도를 함께 이용할 수 있다면, 사회연결망에서 실제로 영향력이 큰 사용자를 올바르게 파악할 수 있을 것이다. 본 논문에서는 블로그 연결망 내에 사용자들간의 관계의 정도를 파악할 수 있는 정보가 저장되어 있다는 점에 착안하여 이러한 관계 정도를 함께 분석하여 실제로 블로그 서비스를 활성화하는데 공헌하는 파워 유저를 파악할 수 있는 새로운 방안을 제시한다.

2. 연구동기

블로그(blog)는 사용자가 자신의 글을 올릴 수 있는 일종의 개인 웹사이트이다[4][5][6]. 블로그 연결망 내의 각 사용자가 자신의 블로그에 게시글을 작성하면, 다른 사용자들은 이 게시글에 대하여 (1) 조회하거나, (2) 댓글을 남기거나, (3) 스크랩하거나, (4) 엮인글 달기를 할 수 있다. 여기서, 조회하기는 게시글을 읽는 액션이며, 댓글 남기기는 게시글에 대하여 자신의 의견을 남기는 액션이다. 또한, 스크랩 하기는 게시글의 내용을 복사하여 자신의 블로그 내에 게시글로 등록하는 액션이며, 엮인글 달기는 게시글과 연관된 새로운 내용을 자신의 블로그 내에 게시글로 작성하는 액션이다. 스크랩하기와 엮인글 달기를 통하여 재생산된 게시글도 일반 게시글과 마찬가지로 조회하기, 댓글 남기기, 스크랩하기, 엮인글 달기 등의 새로운 액션을 유발할 수 있다.

블로그 연결망 전체에 영향을 미치는 소수의 파워 유저

를 파악할 수 있다면, 이들만을 대상으로 블로그 연결망 활성화를 위한 경제적인 정책을 수행할 수 있다. 예를 들어, 파워 유저들이 활발하게 블로그 서비스를 이용하도록 장려함으로써 전체 블로그 연결망의 활성화를 위한 촉매제의 역할을 수행하도록 할 수 있다.

블로그 연결망은 기존의 온라인 연결망과 달리 사용자들간의 관계에 게시글 내의 콘텐츠가 중요한 매개체의 역할을 담당하고 있다. 본 논문에서는 이 점에 착안하여 블로그 연결망 내에서 콘텐츠의 질을 분석함으로써 다른 사용자들의 액션에 크게 영향을 미치는 사용자들을 파악하는 방안에 대하여 논의한다.

3. 제안하는 기법

앞으로의 논의 전개를 위해서 필요한 용어 및 기호들을 정리하면 <표 1>과 같다.

<표 1> 용어 정의.

U_i : 식별자가 i 인 사용자
$D_i = \{D_{i,1}, D_{i,2}, \dots\}$: 사용자 U_i 가 소유한 게시글들의 집합
D_{ij} : U_i 의 블로그에 등록된 j 번째 게시글
$DocumentContentsPower(D_{ij})$: 게시글 D_{ij} 의 콘텐츠 파워
$UserContentsPower(U_i)$: 사용자 U_i 의 콘텐츠 파워
$ActionType = \{W, R, C, S, L\}$: 블로그 서비스를 이용한 사용자 액션
$ActionWeight = \{w_w, w_r, w_c, w_s, w_l\}$: 사용자 액션의 가중치

특정 게시글에 대하여 다른 사용자들이 액션을 보인다는 것은 이 게시글로 인하여 영향을 받았음을 의미한다. 이와 같은 사실에 착안하여 본 연구에서는 게시글의 콘텐츠 파워를 계량화하기 위해서 각 게시글에 대한 사용자들

의 액션의 가중치와 빈도를 곱하는 방법을 사용한다. <표 2>는 게시글의 콘텐츠 파워의 계량화 방법을 정리한 것이다.

<표 2> 게시글의 콘텐츠 파워 계산.

$$\begin{aligned} & \text{DocumentContentsPower}(D_{ij}) \\ &= w_D * \text{DirectContentsPower}(D_{ij}) + w_I * \text{IndirectContentsPower}(D_{ij}) \\ & \text{DirectContentsPower}(D_{ij}) \\ &= w_r * R_Count(D_{ij}) + w_c * C_Count(D_{ij}) + w_s * S_Count(D_{ij}) \\ & \quad + w_l * L_Count(D_{ij}) \\ & \text{IndirectContentsPower}(D_{ij}) \\ &= w_D * \sum \text{DirectContentsPower}(D_{i'j'}) + w_I \\ & \quad * \sum \text{IndirectContentsPower}(D_{i'j'}) \end{aligned}$$

여기서, $D_{i'j'}$ 은 D_{ij} 로부터 직접적으로 재생산된 게시글을 의미

사용자의 콘텐츠 파워는 사용자 블로그에 등록되어 있는 모든 게시글들의 콘텐츠 파워를 이용하여 계산한다. 게시글의 콘텐츠 파워는 노출된 시간에 비례하여 증가하는 경향이 있다. 따라서 오래전에 등록된 게시글은 최근에 등록된 게시글에 비하여 실질적인 콘텐츠의 영향력은 작더라도 오랜 노출 시간(exposed time) 때문에 큰 콘텐츠 파워를 가지는 것으로 왜곡될 수 있다. 본 연구에서는 이러한 문제를 해결하기 위해서 각 게시글의 콘텐츠 파워에 등록 이후의 노출 시간을 반영하는 방법을 사용한다. 즉, 전체 블로그 데이터의 분석 기간을 1로 간주하고, 해당 게시글이 등록된 시각을 참조하여 해당 게시글의 상대적 노출 시간을 계산한다. 이 노출 시간의 역수를 게시글의 콘텐츠 파워에 곱함으로써 노출 시간의 차이로 인한 왜곡을 보정한다.

3. 블로그 연결망 분석

본 실험에서 성능 평가의 대상으로 선정된 파워 유저 선정 기법은 연결 중앙성을 이용하는 기법(DEG), 사용자가 직접적으로 생산한 게시글들만을 대상으로 하는 콘텐츠 파워를 이용하는 기법인 ContentsPower_Direct(CP_D), 사용자가 스크랩하기 혹은 워인글 달기 등의 액션을 통하여 간접적으로 생산한 게시글들만을 대상으로 하는 콘텐츠 파워를 이용하는 기법인 ContentsPower_Indirect(CP_I), 직접적으로 생산한 게시글들과 간접적으로 생산한 게시글들을 모두 대상으로 하는 콘텐츠 파워를 이용하는 기법인 ContentsPower_Total(CP_T)을 사용하였다.

(그림 1)은 실험 결과를 나타낸다. 실험 결과, 기존의 위상구조만을 이용하여 파워 유저를 서정한 기법인 DEG에 의하여 선정된 파워 유저 중에서 사용자 ID 307819, 사용자 ID 227694, 사용자 ID 191454, 사용자 ID 1050536 등 넷 외에는 다른 기법에 의하여 선정된 파워 유저 집합에 전혀 포함되지 않은 것으로 나타났다. 이러한 결과는 사용자간의 관계의 정도에 관한 정보를 분석하여 파워 유저를 선정하는 기법들의 결과는 위상 구조만을 이용하여 파워 유저를 선정하는 기법의 결과와는 매우 다르다는 것

을 의미한다. 따라서 위상 구조만을 이용하는 기존의 DEG는 블로그 연결망에서 큰 영향력을 가지는 파워 유저를 찾기 위한 올바른 방법이 될 수 없다. 파워 유저 선정 시에는 응용에서 요구하는 특성을 만족하는 기법을 사용해야 한다.

rank	DEG	Contents Power		
		CP_D	CP_I	CP_T
1	21608	307819	205649	307819
2	1212576	227694	521114	227694
3	1223870	360763	1066638	521114
4	307819	957182	609211	360763
5	173326	240513	38291	957182
6	571678	976758	228000	205649
7	379712	1007110	156438	976758
8	360430	275612	579543	240513
9	720669	1329246	1283112	1007110
10	1030582	1254859	29155	1153439
11	221718	1153439	481166	275612
12	227694	1050536	330592	1329246
13	1269927	954535	1139729	1254859
14	465355	1214458	1191046	1050536
15	191454	84535	591032	1214458
16	1138925	1120363	660655	954535
17	1263255	1041565	1336501	579543
18	1244929	782455	1223256	999317
19	108545	208456	1384606	84535
20	127904	999317	938203	1120363
21	1050536	493135	975704	782455
22	351335	191454	51090	1041565
23	985608	140607	1009253	493135
24	629130	417359	33801	140607
25	608703	371450	1380719	191454
26	709532	1362815	1354168	208456
27	561782	337868	745863	84463
28	433300	1049339	269898	417359
29	639	1341008	1032290	371450
30	336602	84463	513037	966173

(그림 1) 각 기법에 따른 파워 유저 집합의 결과.

4. 결론

본 논문에서는 블로그 연결망에서 파워 유저를 파악하는 방안에 대하여 논의하였다. 기존의 연구에서는 사회연결망의 관계 정도에 대한 정보의 부재로 인해서 위상 구조적인 특징만을 고려하여 파워 유저를 파악하였다. 그러나 이러한 방식은 응용의 요구 사항을 원천적으로 배제하므로 대부분의 응용에 적합하지 않다. 본 논문에서는 블로그 연결망 내에 사용자들간의 관계의 정도를 파악할 수 있는 정보가 저장되어 있다는 점에 착안하여 이러한 관계 정도를 함께 분석함으로써 블로그 서비스를 활성화하는데 실제로 공헌하는 파워 유저를 파악하는 새로운 방안을 제시하였다.

감사의 글

본 연구는 NHN(주)의 지원을 받았습니니다. 그러나, 본 논문에서 제시된 의견이나 결론, 또는 권고 등은 온전히 저자(들)의 것이며, 반드시 지원회사의 입장을 대변하는 것은 아닙니다.

참고문헌

[1] L. Adamic, O. Buyukkokten, and E. Adar, "A Social Network Caught in the Web," First Monday, Vol. 8, No. 6, pp. 1-22, 2003.

[2] R. Kumar, J. Novak, and A Tomkins, "Structure and Evolution of Online Social Networks," In Proc. Int'l. Conf. on Knowledge Discovery and Data, pp. 611-617, 2006.

[3] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the Spread of Influence through a Social Network," In Proc. ACM Int'l. Conf. on Knowledge Discovery and Data Mining, ACM SIGKDD, pp. 137-146, 2003.

[4] (주)SK Communications, <http://www.cyworld.com>.

[5] (주)엠포스, <http://blog.empas.com>.

[6] (주)NHN, 블로그홈, blog.naver.com.